Renewable and Sustainable Energy Reviews 155 (2022) 111909

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Ensemble solar forecasting and post-processing using dropout neural network and information from neighboring satellite pixels

Gokhan Mert Yagli [a], Dazhi Yang [b,*], Dipti Srinivasan [c]

[a] Solar Energy Research Institute of Singapore (SERIS), National University of Singapore (NUS), Singapore
[b] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, Heilongjiang, China
[c] Department of Electrical and Computer Engineering, National University of Singapore (NUS), Singapore

ARTICLE INFO

Keywords:
Dropout neural network
Ensemble solar forecasting
Machine learning
Monte Carlo sampling
Post-processing
Satellite-derived irradiance

ABSTRACT

Ensemble weather forecasts are often found to be under-dispersed and biased. Post-processing using spatio-temporal information is, therefore, required if one wishes to improve the quality of the raw forecasts. It is on this account that the present article generates and post-processes ensemble solar forecasts using satellite-derived irradiance not only from the focal pixel but also from the neighboring pixels. The ensemble forecasting model of choice is a dropout neural network with Monte Carlo sampling, eliminating the need for training multiple models and ensuring parameter diversity in ensemble forecasting. Subsequently, ensemble forecasts are post-processed using both parametric and nonparametric post-processing techniques, such as nonhomogenous regression, generalized additive model, linear quantile regression, or quantile random forests. The proposed forecasting framework is demonstrated and verified using four years of half-hourly data, at seven locations in the United States. Continuous ranked probability skill scores as high as 66% have been obtained when comparing the proposed method to a conditional climatology reference. The content of this article may be useful to a wide range of stakeholders in the power system, including but not limited to: independent system operators, who aim at efficiently maintaining the system's reliability; utility- and distributed-scale PV plant owners, who wish to avoid penalties for power deviation between the scheduled and real-time delivery; and forecast retailers, who can benefit from selling solar forecasts of higher quality.

1. Introduction

Solar forecasting is thought to be vital to power grid operations under high renewable penetration, but most people seem to have demonstrated what is known as the illusion of explanatory depth in regard to this subject. One such fact in support of our proposition is related to ensemble forecasting, or forecast combination, as also commonly known. Though the benefit of ensemble forecasting is one of the few things on which professional forecasters have a consensus [1], its utilization has hitherto been limited in solar forecasting [2]. In contrast, the bulk of the current solar forecasting literature is on deterministic forecasting, which is unable to quantify the uncertainty associated with the forecasts, and thus limits power system engineers' confidence in maintaining the stability of the power system against the disturbances caused by the variable solar power generation [3]. The chief property of ensemble forecasting is that it requires not just a single "best-guess" forecast, but rather multiple *component forecasts*, or member forecasts, so as to mitigate the bias in the judgment of a single forecaster, to improve forecast quality, and to quantify the uncertainty [4,5].

That said, ensemble forecasting is not problem-free. As noted in several recent works, ensemble solar forecasts are often under-dispersed [6], biased [7,8], and aggregate inconsistent [9–11]. Thus, it is clearly necessary to calibrate, or post-process, the raw ensemble forecasts to improve their quality. Well-known calibration methods include Bayesian model averaging, generalized additive models (GAMs), and ensemble model output statistic, or nonhomogenous regression (NR).

Abbreviations: BON, Bondville; CRPS, continuous ranked probability score; CSI, clear-sky index; DRA, Desert Rock; FPK, Fort Peck; GAM, generalized additive model; GHI, global horizontal irradiance; GWN, Goodwin Creek; LQR, linear quantile regression; MC, Monte Carlo; MLE, maximum likelihood estimation; NGR, nonhomogenous Gaussian regression; NN, neural network; NR, nonhomogenous regression; NSRDB, National Solar Radiation Data Base; PICP, prediction interval coverage probability; PIT, probability integral transform; PS, preselected; PSU, Pennsylvania State University; QRF, quantile random forests; SURFRAD, Surface Radiation Budget Network; SXF, Sioux Falls; TBL, Table Mountain; trLO, truncated logistic distribution; trSST, truncated skewed Student's t distribution; VAR, vector autoregression

* Corresponding author.
E-mail address: yangdazhi.nus@gmail.com (D. Yang).

https://doi.org/10.1016/j.rser.2021.111909
Received 22 July 2021; Received in revised form 12 October 2021; Accepted 9 November 2021
Available online 1 December 2021

The reader is referred to [5,12] for reviews on these and other calibration methods.

Besides calibration, another factor that heavily impacts the final ensemble forecasting results is the quality of component forecasts [12]. One must choose the component forecasting method adequately, with a careful integration of domain knowledge. One of the salient features of solar irradiance is its spatio-temporal nature, which can often be captured through the means of remote sensing—data from geostationary, polar-orbiting, and deep-space satellites [13]. Because variability in solar irradiance is induced mostly by moving clouds, with no surprise, spatio-temporal solar forecasting approaches, which are able to capture such dynamics of clouds, have been shown to be more advantageous than forecasting approaches that only use local information [14–16]. In what follows, we motivate our proposal, that is, using data from neighboring satellite pixels to perform ensemble forecasting and post-processing.

1.1. Generating component forecasts

Ensemble solar forecasts can be generated by either physics-based (e.g., numerical weather prediction with perturbed initial conditions) or data-driven (e.g., through a collection of statistical and machine-learning models) approaches [17]. In both cases, to reflect the forecast uncertainty adequately, diversification in forecasts of various ensemble members is an important concern. Diversification can be created using different data, models, and parameter sets. When the same forecasting model is trained using different datasets, the forecasts are naturally different [18]. Similarly, one can assume several data-generating processes, and thus builds several models using the same dataset, each explaining one probable data-generating process [19,20].

The last type of diversification involves using different model-parameter sets to create different versions of forecasts. In the case of linear models with many predictors, once the full model is fitted, one can randomly select some regression coefficients and set them to zero, so that the prediction is made without the contributions (or zero weights) from those variables. If this step is repeated many times, a set of predictions can be obtained, which gives a probabilistic representation of the prediction. In the case of nonlinear models, e.g., neural networks (NNs), once the NN model is trained, predictions are repeatedly made by randomly omitting some nodes/links each time. Such an NN model is known as a *dropout NN* [21], which has been recently considered in a solar forecasting context, see e.g., [22–24]. Note that dropout NN is not to be confused with the dropout technique used for regularization to prevent over-fitting of NN [25]. Indeed, the dropout NN is used to generate the raw ensemble forecasts in this article. It should be noted that the traditional dropout NNs only apply dropout during training, and the forecasts/predictions during testing are deterministic. The current dropout NN gains the capability of generating probabilistic forecasts by applying Monte Carlo (MC) dropout during testing.

The reason for opting for the dropout NN in this work needs to be discussed in concert with the two main traditional classes of methods for generating probabilistic forecasts using satellite-derived irradiance [26]. The first class uses optical flow or cross-correlation method to determine the motion of clouds, and thus convert the advected cloud field to irradiance. To generate probabilistic forecasts in this scenario, method of dressing is required [5], which adds complexity to the forecasting problem. Dressing assumes that past errors of a forecasting model, under a particular forecasting condition, can be used for the current forecast from the same model and under the same condition, which may be reasonable to quantify the uncertainty associated with forecasts [27]. On top of that, the cloud-to-irradiance conversion step is seen as a major barrier preventing accurate forecasting, due to difficulties such as the parallax effect during ray tracing [28], or orographic effect over complex terrain [29]. To that end, using an NN could substantially reduce the amount of physical modeling during forecasting, which may be more amenable to forecast practitioners [30,31].

The second class of methods uses spatio-temporal statistical models, such as vector autoregression (VAR) or kriging, to generate forecasts using time series of spatial processes [32,33]. Although these methods can generate probabilistic forecasts by default, they can be quite problematic when the dimension of the input becomes large [34–36]. This is particularly true when the spatial resolution of gridded irradiance data is becoming increasingly higher. In this regard, deep learning approaches are known to be capable of handling big data, and their scalability is often reasonable.

1.2. Ensemble forecast post-processing

As mentioned earlier, ensemble forecasts are often not calibrated, that is, the predictive distributions can be too narrow (over-confident) or too wide (under-confident) [12]. Hence, it is a good practice to perform diagnosis and post-process the raw ensemble forecasts whenever needed. In view of our previous discussions on the importance of utilizing spatio-temporal information during forecasting, it is useful to distinguish spatial post-processing from conventional post-processing.

Conventional post-processing exploits the correspondence between forecasts and observations at a single location, which we call *focal location*. Taking NR for an example, the parameters of the final predictive distributions (e.g., mean and variance) are modeled as regression functions of component forecasts. Then, by minimizing a loss function or maximizing the likelihood, regression coefficients, which act as weights as to combine the component forecasts, are found and used to post-process subsequent forecasts. Nonetheless, this approach ignores the useful information embedded in forecasts at neighboring locations [37,38]. For instance, the aforementioned parallax effect may lead to a scenario where a better forecast for the focal location falls into an adjacent pixel, which in turn can be used as a predictor in NR. To that end, spatial post-processing not only uses forecasts at the focal location but also includes forecasts at its surroundings during training.

In fact, in the forecasting of other weather variables, such as temperature, precipitation, or wind speed, post-processing techniques leveraging regional forecasts have been studied and shown effective in some cases, e.g., [39–41]. In contrast, post-processing studies in the solar forecasting literature have largely been limited to local information [5], with notable exceptions [10,11]. Hence, in this article, the main contribution is that the potential benefits offered by incorporating spatio-temporal information into ensemble post-processing are formally investigated in a solar forecasting context.

The remaining part of the article is organized as follows. Section 2 describes data from the Surface Radiation Budget Network (SURFRAD) and the National Solar Radiation Data Base (NSRDB). Section 3 describes the forecast generation and post-processing methodology. The results are presented in Section 4. Conclusions follow in Section 5.

2. Data

A reliable data source is essential to solar engineering applications. In this study, two such high-quality datasets are considered, one is ground-based, and the other is satellite-derived.

2.1. Ground-based data, SURFRAD

SURFRAD is a network of seven research-grade monitoring stations covering major climate zones located throughout the contiguous United States [42]. Global horizontal irradiance (GHI) measurements in 1-min resolution over 2015–2018 are downloaded. The raw measurements are quality controlled and aggregated to 30-min resolution using `SolarData` package available in the R programming language [43,44]. We followed the procedures described in [43–45] for GHI data quality control and preprocessing steps. All seven SURFRAD stations are used in this study, namely, Bondville (BON), Desert Rock (DRA), Fort Peck (FPK), Goodwin Creek (GWN), Pennsylvania State University (PSU), Sioux Falls (SXF), and Table Mountain (TBL).

2.2. Satellite-derived data, NSRDB

Satellite-derived irradiance is often taken as a fairly reasonable approximation of surface irradiance, in situations where ground-based radiometric data is unavailable [46–48]. However, satellite-derived irradiance is prone to systematic errors and under-dispersion [49]. Hence, such data should be validated, and corrected if necessary, before any further application [48]. NSRDB satellite-derived irradiance data were validated against SURFRAD data, and it is found that NSRDB irradiance data agree well with SURFRAD's observations—hourly GHI normalized root mean square error is reported as in 8.9–18.7% [50]. Hence, in this study, satellite-derived irradiance values are sourced from NSRDB.

NSRDB contains GHI estimates and their corresponding clear-sky estimates at a 4 × 4 km spatial resolution and a 30-min temporal resolution for most of the Americas [51]. Clear-sky GHI expectations available in NSRDB are computed using REST2 model [52], which is identified as one of the most accurate clear-sky models in the literature [53,54], and is well-suited for solar forecasting [55,56]. The GHI values and their corresponding clear-sky expectations over 2015–2018 are downloaded for 81 satellite pixels surrounding each SURFRAD station. These 81 pixels roughly cover an area of 3378.00–4184.98 km^2. In a more general sense, one can change the number of pixels to track cloud movements in spatio-temporal forecasting, considering, e.g., dominant cloud type and wind profile over a focal station, but that is not considered in this paper.

3. Methodology

Before exploring further methodological details, an overview of the computer experiment is given for clarity and ease of reader. In this work, 30-min-ahead data-driven ensemble clear-sky index (CSI) forecasts are generated using a dropout NN model using data from the SURFRAD ground stations and the NSRDB satellite-derived irradiance. The reason why CSI is used as input instead of irradiance is that the double-seasonal pattern in solar irradiance is primarily due to the apparent Sun–Earth movement and atmospheric transmittance, which can be estimated quite accurately with a clear-sky model. Indeed, removing the seasonal component from time series is a well-accepted forecasting practice [57]. Nevertheless, CSI forecasts are back-transformed to irradiance, by multiplying the clear-sky irradiance at the forecast time stamps, before verification. Stated differently, the final error calculations and visualizations are still performed in irradiance terms.

Once the raw CSI forecasts are generated, these ensemble CSI forecasts are post-processed using (1) forecasts from the focal location only (referred to as *focal post-processing* hereafter), and (2) forecasts from focal and surrounding pixels (referred to as *spatial post-processing* hereafter). More specifically, the post-processing step takes a regression form, in which the regressors are forecasts from focal pixel (case 1), or focal and surrounding pixels (case 2), whereas the regressand is the actual CSI observed from the focal location. These two case studies are to contrast whether or not leveraging neighboring information is helpful in improving the quality of the post-processed forecasts.

The following steps are taken for ensemble forecast generation and post-processing. Ensemble forecasts for each SURFRAD station and 81 surrounding satellite pixels around these stations available at NSRDB are first generated. Secondly, ensemble forecasts at each SURFRAD station are post-processed. A preselection approach is proposed to improve the computational effectiveness of spatial post-processing. Finally, the performance of raw ensemble forecasts and post-processing techniques in focal and spatial post-processing are subsequently compared. Fig. 1 depicts a schematic diagram of the main steps taken for forecast generation, post-processing and verification.

3.1. Ensemble forecast generation, Monte Carlo dropout

Dropout is a stochastic regularization technique that prevents NNs from overfitting [25]. Dropout randomly deactivates neurons during training and can be implemented in all layers except for the output layer. There are different implementations of dropout in popular deep-learning frameworks, the default `Keras` implementation is used here. In `Keras` dropout implementation, for example, suppose dropout is used with rate $p = 0.2$ at a layer of NN having 100 neurons, randomly chosen 20% of neurons are temporally deactivated along with their incoming and outgoing connections, and thus 80 neurons are used in forward and backward passes. Subsequently, a new set of 20% of the neurons in that particular layer is randomly picked and deactivated again in the next training case.

At this stage, it is important to remark that the original proposal of the dropout technique only applies during the training phase of NNs. In other words, no neurons are deactivated during testing phase, i.e., ensemble forecast generation in our case. Hence, the original proposal of the dropout technique alone does not explicitly reflect the uncertainty associated with the forecasts. Since uncertainty information is critical for operational energy forecasting, dropout NN models should be modified to reflect forecast uncertainty.

A theoretically valid and computationally feasible way to obtain forecast uncertainty information from a dropout NN model is shown in [21]. In this approach, dropout is applied at each time t in testing phase—unlike the original implementation of the dropout technique—which creates a posterior distribution. Subsequently, MC samples are extracted from this distribution. This procedure is referred to as the *MC dropout*. Extracted MC samples are taken as ensemble forecasts, and thus permits one to obtain a predictive distribution from a single NN model. Since the complete proof is lengthy with in-depth discussions on Bayesian modeling and Gaussian processes and not needed in what follows in this article, the interested reader is referred to [21] for details.

The hyperparameters of NN models have a direct implication on the predictive performance. For this reason, model users often put in much effort in tuning hyperparameters, which is undoubtedly a tedious process relying largely on computationally prohibitive empirical approaches. In dropout NN, dropout rate p is a hyperparameter that should be carefully tuned for being effective. However, a practical variant of the traditional dropout technique, called *concrete dropout*, allows dropout rate p to adapt and to be optimized using gradient methods and pathwise derivative estimator by minimizing an objective function at each layer where dropout is implemented [58]. Since the mathematical derivation of concrete dropout is lengthy, interested readers are referred to [58]. The Python implementation of concrete dropout and the authors' recommendations for hyperparameter initialization are available in `yaringal/ConcreteDropout` GitHub repository, and R implementation is given in [59], which are followed for the basis of the work presented in this article.

Besides dropout rate, there are numerous hyperparameters to be tuned in NNs, for example, the number of neurons/layers, activation functions, weight initialization, or learning rate [60]. However, in this work, we refrain from tuning hyperparameters—except for the dropout rate that is internally tuned during training—for two reasons. The first reason is this: the dropout NN model needs to be individually trained for each of 81 satellite pixels and 1 ground station to generate spatial forecasts around a location of interest. As a result of the individual hyperparameter tuning process, a different optimum hyperparameter set would most likely be found for each NN model. Also, the method comparison needs to be repeated for 7 SURFRAD stations, which requires tuning of 82 × 7 NN models. Individually tuning such a large number of NN models, even over a minimal hyperparameter search space, would not be technically efficient. The second reason is this: the aim of this article is to introduce the new techniques for ensemble solar forecasting, nor is there a need to put too much emphasis on what is

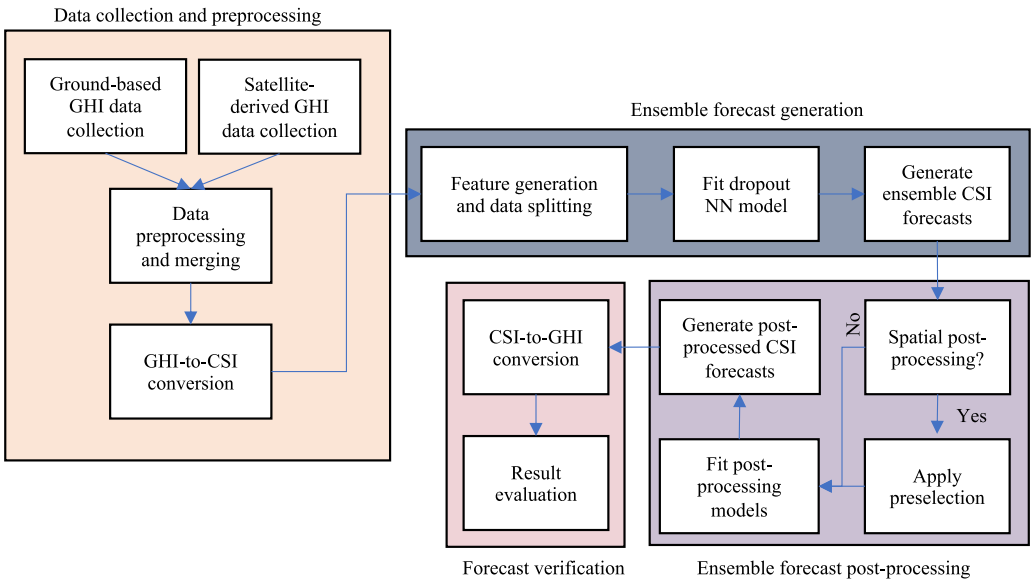

Fig. 1. The schematic of the steps taken for ensemble forecast generation, post-processing and verification in this study.

best for the particular dataset at hand, since the real difficulty is *always* concerned with those datasets that have yet been tested—this issue is more generally known as the *problem of induction*, as in the philosophy of science.

To summarize, the input layer has 82 neurons (for 81 satellite pixels + 1 ground station). The output layer consists of two neurons: one predicting the conditional mean, and the other predicting the logarithm of variance. The NN output layer with two neurons is a common structure seen in probabilistic forecasting studies if a two-parameter statistical distribution, e.g., Gaussian, is assumed to model the predictive distribution [41]. As opposed to the common loss functions, such as the mean square error as commonly adopted for deterministic forecasting with single-output NN, two-output NN used in this article optimizes a heteroscedastic loss function [61], which is formulated as follows:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{t}\frac{1}{2\hat{\sigma}_t^2}\left(y_t - \hat{y}_t\right)^2 + \frac{1}{2}\log\hat{\sigma}_t^2, \tag{1}$$

where t indexes time; the parameters to be optimized are denoted as θ; N is training data length; $\hat{\sigma}_t^2$ is the predicted variance at t; and y_t and $\hat{y}_t$ are the actual observation and the predicted mean at t, respectively. In this study, $\mathcal{L}(\theta)$ is optimized using the Adam optimizer [62].

3.2. Forecasting setup

In this work, two years (2015 and 2016) of 30-min CSI data are used to train the NN model, and 30-min-ahead ensemble CSI forecasts are generated for the remaining two years (2017 and 2018), again, in 30-min resolution. The forecast horizon and data resolution are both 30 min. More specifically, 30-min-ahead CSI forecasts are generated using 30-min CSI observations from the nearby satellite pixels and the ground station up to forecast time. For example, the CSI value at 14:30 UTC is predicted using the CSI values from the ground station and satellite pixels up to 14:00 UTC. Similarly, multi-step-ahead forecasts can be generated by slightly modifying this setting. For example, 1-h-ahead forecast for $t + 1$ h can be generated using CSI observations up to time t, where the response of training data corresponds to the observation at time $t + 1$ h.

The left panel of Fig. 2 illustrates the grid of satellite pixels around the DRA station, in Desert Rock, Nevada. Whereas the DRA ground station is shown in red, its surrounding satellite pixel centers are depicted in blue. Using the aforementioned dropout NN model, 50 component forecasts are generated for each time instance at the DRA station. These 50-member ensemble forecasts are post-processed using focal post-processing approaches. However, a central theme of this article has been advocating the use of forecast information at neighboring pixels during post-processing. To that end, component forecasts are generated not only at the focal station itself but also at the neighboring pixels; the right panel of Fig. 2 shows one such scenario, where forecasts for the red satellite pixel are generated based on the blue satellite pixels and the DRA ground station. In spatial solar forecasting, the number of component members is set to 20 for the neighboring satellite pixels and the ground station as to regulate computational burden.

For spatial solar forecasting, after the forecast-generation procedure is performed at the focal station and is repeated for all 81 pixels, one obtains a total of $82 \times 20 = 1640$ component forecasts, for each time instance. Clearly, the number of ensemble members is abundant, and a potential problem of information redundancy arises. Therefore, a member-selection step is thought to be necessary. In short, we reduce the initial set of 1640 component forecasts to a set of 50, based on some selection criteria that will be discussed in Section 3.3.3. The number of ensemble members is set to 50 arbitrarily, and perturbing this number to be reasonably bigger or smaller will have a negligible effect on the final results. These 50-member ensemble forecasts are post-processed using spatial post-processing techniques.

3.3. Post-processing of ensemble forecasts

Ensemble CSI forecasts generated by the MC dropout model are post-processed using state-of-the-art parametric and nonparametric techniques. The NR-/GAM-based techniques with 2- and 4-parameter distributions are employed for parametric post-processing, whereas nonparametric techniques are based on quantile regression. Next, these post-processing techniques and data splitting exercise for post-processing are elaborated.

3.3.1. Parametric post-processing

One of the most popular parametric post-processing techniques used in meteorology is the nonhomogeneous Gaussian regression (NGR), in that, the shape of the distribution of temperature ensemble forecasts for each time t is assumed to be normal [63]. In NGR, the predictive distribution of post-processed forecast $\tilde{y}$ at time t is assumed to distribute according to:

$$\tilde{y}_t \sim \mathcal{N}\left(\mu_t, \sigma_t^2\right), \tag{2}$$

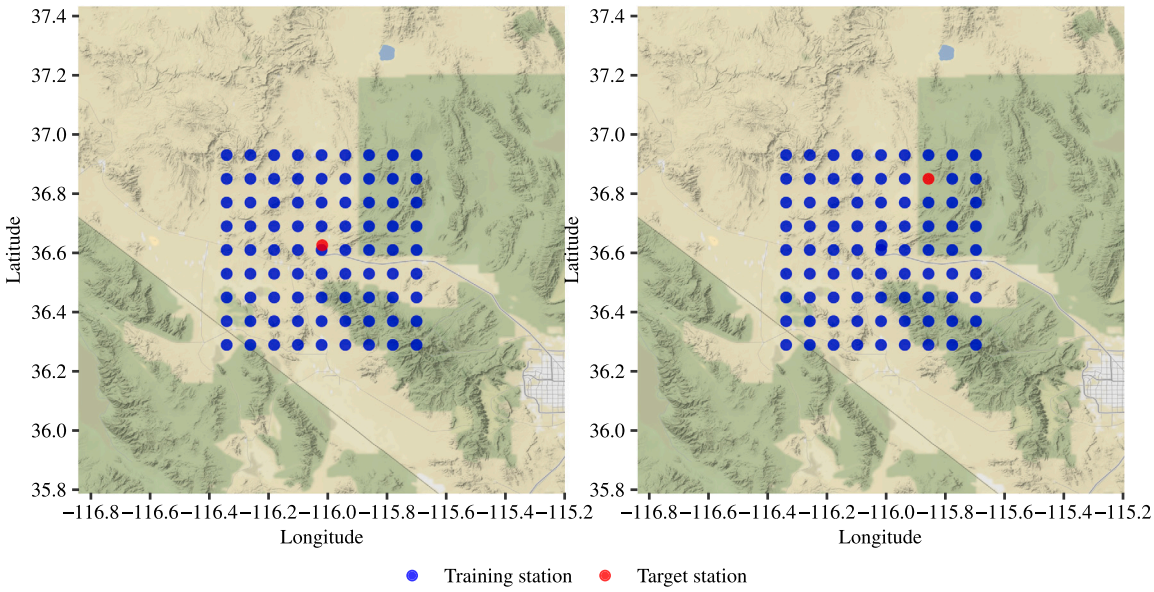

Fig. 2. Satellite pixels around the Desert Rock ground station, see Section 3.2 for interpretation as it is too complex to be fit into the caption.

where μ_t and σ_t^2 are the mean and variance of a Gaussian distribution, which are,

$$\mu_t = b_0 + b_1\hat{y}_{t,1} + b_2\hat{y}_{t,2} + \cdots + b_M\hat{y}_{t,M}, \tag{3}$$

$$\sigma_t^2 = c + dS_t^2, \tag{4}$$

where $\hat{y}_{t,k}$ is the kth ensemble component forecast at time t; b_0 is the intercept; b_k is the post-processing weight (or mixing coefficient) assigned to the kth ensemble component forecast; c and d are the assigned coefficients to modify S_t^2; and S_t^2 is the *ensemble variance* calculated from the component members:

$$S_t^2 = \frac{1}{M}\sum_{m=1}^{M}\left(\hat{y}_{t,m} - \bar{y}_t\right)^2, \tag{5}$$

where $\bar{y}_t$ is the mean of all component forecasts at time t. One should note that $M = 50$ in this case. By using historical observation–forecast pairs, the coefficients $b_0, \ldots, b_M$, c, and d in the Eqs. (3) and (4) can be estimated.

NGR is only appropriate if the ensemble forecasts exhibit a Gaussian distribution. However, that is rarely the case for atmospheric variables. If the forecast variable violates the imposed distribution assumptions, alternative predictive distributions should be considered. On this point, the original NGR formulation, namely, Eqs. (3)–(5) can be further generalized by assuming different predictive distributions instead of Gaussian distribution. For instance, [64] assumed the wind speed ensemble forecasts to follow a gamma distribution.

More general than NGR is the GAM-based statistical post-processing technique, which was introduced to solar forecasting very recently [6,8]. Before elaborating GAM-based post-processing, unimodal statistical distributions and their parameters should be explained. Four parameters jointly define the shape of a unimodal distribution: location (μ), scale (σ), skewness (ν), and kurtosis (τ). The "center" and the "spread" of the distribution are generally defined with the location and scale parameters, respectively. Skewness and kurtosis parameters measure—relative to the normal distribution—the asymmetry and tailedness of the distribution [65]. For instance, the logistic distribution has 2 parameters, location and scale.

The mean of a Gaussian distribution is modeled using a linear combination of component forecasts as in Eq. (3). Such linear modeling can be further generalized to model all parameters of a statistical distribution in GAM-based post-processing. The parameters at each forecast time t can be formulated as follows:

$$\mu_t = b_0 + \sum_{m=1}^{M} b_m\hat{y}_{t,m}, \tag{6}$$

$$\sigma_t = c_0 + \sum_{m=1}^{M} c_m\hat{y}_{t,m}, \tag{7}$$

$$\nu_t = d_0 + \sum_{m=1}^{M} d_m\hat{y}_{t,m}, \tag{8}$$

$$\tau_t = e_0 + \sum_{m=1}^{M} e_m\hat{y}_{t,m}, \tag{9}$$

where b_0, c_0, d_0, e_0 are the intercepts, and b_k, c_k, d_k, and e_k are the coefficients assigned to the kth ensemble member at time t.

The intercepts and coefficients in Eqs. (6)–(9) need to be estimated in order to construct the predictive distributions. To estimate the corresponding regression coefficients, two common approaches are used in the literature: (1) by maximizing a likelihood function (also known as maximum likelihood estimation, or MLE), and (2) by minimizing a loss function—the most commonly used loss function is the continuous ranked probability score (CRPS). Both estimation techniques lead to very similar results if the distribution assumption is suitable [66]. In this study, MLE routine is adopted for fitting ensemble parameters as the analytical forms of CRPS for many parametric distributions are not readily available and require derivation, which hinder the applicability of CRPS minimization. Given N observations, y_t, $t = 1, \ldots, N$, and their corresponding probability density functions, $f^{\hat{y}_t}(y;\theta)$, the log-likelihood $\ell(\theta) = \sum_{t=1}^{N} \log f^{\hat{y}_t}(y_t;\theta)$ is maximized to estimate the parameters.

In GAM-based techniques, the response variable is assumed to follow a statistical distribution, and the parameters of the distribution vary according to ensemble forecasts for each time t [67]. GAM-based post-processing works similar to NR-based models, but the regression framework differs, see [68] for details. GAM-based post-processing is often found slightly superior to NR-based post-processing due to its versatility [6].

In regard to parametric post-processing techniques, two important aspects stand out. One of them is the choice of statistical distribution, which plays a crucial role in attaining good accuracy of post-processing. The other is that truncation of the chosen distributions on the x-axis from the left side, the right side, or both sides, as to constrain non-physical results [65]. For example, although CSI is non-negative, negative CSI values may appear as lower quantiles

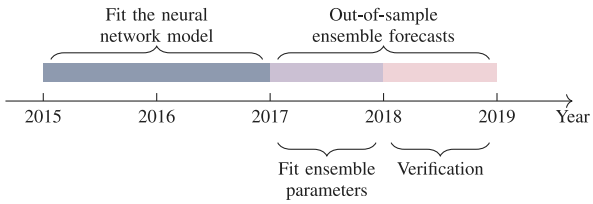

Fig. 3. Train–test split exercise used in this article for forecasting and post-processing.

after post-processing. Hence, distributions should be truncated to reflect such constraints. Because the predictive performance of several truncated distributions and regression frameworks for CSI ensemble post-processing has been compared in a recent study [6], the current article inherits its choice of statistical distributions and truncation from that article. More specifically, truncated logistic (trLO) and truncated skewed Student's t distributions (trSST) are used, and the distributions are left-truncated at 0 and right-truncated at 1.2.

3.3.2. Nonparametric post-processing

Parametric post-processing techniques allow fast construction of predictive distributions. However, they are often dwarfed in comparison to nonparametric post-processing techniques in terms of performance. The shape of the predictive distribution for nonparametric post-processing techniques is not restricted and it is therefore more flexible. One such technique is linear quantile regression (LQR), which models the relationship between feature variables and specific quantiles of the response variable [69]. LQR is used as a benchmark model owing to its simplicity. Quantile random forests (QRF) proposed by [70] is another commonly used nonparametric post-processing technique, which is based on random forests (RF) [71]. RF generates its forecasts from an ensemble of decision trees, where a random set of features and training samples are used to generate each decision tree. The final prediction is the average of predictions generated by each tree. This framework is extended to QRF, where the full conditional distribution (and therefore quantiles) of the response variable is predicted. QRF is included in this study to exemplify the benefits of incorporating spatio-temporal information into nonparametric post-processing techniques, and to contrast the parametric approaches.

To sum up, the following post-processing techniques are preferred to correct the calibration of CSI ensemble forecasts in this study: (1) NR with truncated logistic distribution (NR+trLO), (2) GAM with truncated logistic distribution (GAM+trLO), (3) GAM with truncated skewed Student's t distribution (GAM+trSST), (4) LQR, and (5) QRF.

3.3.3. Post-processing setup

After applying the procedure outlined in Section 3.2, two years (2017–2018) of out-of-sample ensemble forecasts are obtained. Subsequently, the first year (2017) forecasts are used for the task of parameter estimation for post-processing models (in concert with the corresponding observations, of course), whereas the forecasts from the remaining year (2018) are used for verification. The overall data splitting strategy for both forecasting and post-processing is illustrated in Fig. 3.

Recall that 1640-member ensemble forecasts are generated for each time t to study the effects of spatial forecast information in post-processing. This is problematic on three heads. Take first its most important aspect is that such a huge number of component members may over-represent the uncertainty associated with the forecasts. The second is that the total computation time of post-processing may exceed the forecast horizon, which would render forecasts useless for very short-term forecasting scenarios. Finally, in the preliminary analysis, we noted that GAM-based models cannot converge in some cases with a large number of component forecasts.

One possible solution to these issues, as mentioned earlier, is to reduce the number of ensemble members to be post-processed. For that reason, the ensemble members are preselected before post-processing. At this stage, we apply a correlation-based preselection technique, where the most correlated 50 ensemble members to the response variable, i.e., ground CSI observations, are selected. This is appropriate, as the association between ensemble members and response—both being CSI—is known *a priori* linear. At the end of preselection, 1640 members are reduced to 50 members, on which forecast post-processing is based. To distinguish those ensemble forecasts post-processed with preselection from those without, the prefix of "PS" is used, e.g., PS NR+trLO denotes "preselected ensemble forecasts post-processed by nonhomogeneous regression with truncated logistic distribution."

4. Result and discussion

Goodness of predictive distribution is consisted in its sharpness and calibration. While sharpness denotes the concentration of the predictive distribution, calibration is concerned with the statistical consistency between the distribution of forecast and that of the corresponding observation [72]. Maximal sharpness, subject to being calibrated also, is expected from a perfect predictive distribution. Hence, it is important to evaluate both aspects of goodness.

The quality of predictive distributions is evaluated through both quantitative and qualitative means, where the quantitative metrics are chosen following the suggestions of [73]. The performance statistics calculated at each station are prediction interval coverage probability (PICP), CRPS, and CRPS skill score. PICP assesses calibration for a nominal probability; CRPS is a composite score that assesses calibration and sharpness simultaneously; and CRPS skill score evaluates the improvement over a reference model called the complete-history persistence ensemble is used, which is a conditional climatology tailored to solar forecasting [74]. Also, LQR is used to benchmark other post-processing techniques.

Table 1 shows the results of raw ensemble forecasts, focal and spatial post-processing techniques in terms of 3 quantitative metrics; PICP with a nominal coverage probability of 95%, CRPS, and CRPS skill score. Particularly interesting is that the predictive performance of the MC dropout model significantly outperforms, or is comparable, to the results of 20 fine-tuned data-driven models, as reported in [6], at most of the stations—also recall that the NN is not fine-tuned in this study. Such high predictive performance is in favor of using the MC dropout technique to generate ensemble solar forecasts for spatial forecasting scenarios.

Focal post-processing improves the quality of raw ensemble forecasts, as evidenced by all metrics. All focal post-processing approaches mostly generated predictive distributions having close-to-nominal PICP values, hence indicates better calibration than raw forecasts. Focal post-processing using QRF and GAM+trSST methods mostly generated prediction intervals with the highest quality. The reference LQR model mostly outperformed parametric models with logistic regression (NR+trLO and GAM+trLO). This is expected as nonparametric approaches often outperform their parametric counterparts mainly due to the reasons mentioned in Section 3.3.2. Perhaps of particular note is the good performance of the GAM+trSST model against nonparametric LQR and QRF models in some cases. This emphasizes the importance of proper distribution selection in parametric post-processing.

Using forecast information from neighboring satellite pixels in post-processing provides divergent results in terms of accuracy compared to focal post-processing. In some stations, incorporating forecast information from neighboring satellite pixels provides a clear benefit. However, in some stations, post-processing using focal information only generated higher quality prediction intervals. Whereas the highest CRPS skill scores are in the 44.17–65.85% range for focal post-processing, incorporating spatio-temporal information ended up having CRPS skill scores in the 43.46–65.82% range. Focal post-processing performs as well as

Table 1
Results of focal and spatial post-processing. Row-wise best results are in bold. "PS" stands for preselection, see Section 3.3.3.

Station	Raw	Focal post-processing					Spatial post-processing				
		LQR	NR+trLO	GAM+trLO	GAM+trSST	QRF	PS LQR	PS NR+trLO	PS GAM+trLO	PS GAM+trSST	PS QRF
						PICP [%]					
BON	63.66	93.65	93.33	93.95	94.44	95.75	91.78	93.55	93.10	**95.12**	96.69
DRA	76.53	92.50	88.86	94.25	94.80	**95.20**	94.33	91.96	96.43	95.51	96.79
FPK	52.30	92.78	92.69	**94.86**	93.35	94.38	95.51	94.20	94.11	95.65	96.64
GWN	68.52	92.31	92.38	93.49	93.14	94.00	92.76	93.99	93.50	**95.11**	96.39
PSU	64.19	94.08	93.38	93.34	90.73	93.58	**94.88**	93.76	93.45	94.15	95.52
SXF	65.70	94.89	92.78	93.35	93.75	95.18	95.37	93.96	93.84	**94.99**	96.45
TBL	61.41	91.60	90.51	92.30	92.69	93.55	93.92	93.97	93.66	**94.51**	96.13
						CRPS [W/m^2]					
BON	44.55	39.36	40.32	42.50	37.70	37.60	38.48	39.43	39.96	37.29	**36.84**
DRA	29.47	27.41	30.01	29.61	**26.40**	26.99	27.76	31.19	30.67	27.20	26.74
FPK	50.08	33.06	34.04	36.99	31.91	32.31	33.28	34.45	34.66	31.96	**31.37**
GWN	38.98	36.56	36.89	37.43	37.04	**34.39**	36.25	37.10	37.12	34.97	34.43
PSU	44.42	42.44	42.15	42.18	40.30	39.85	41.61	41.80	42.00	39.85	**39.68**
SXF	37.75	33.17	33.35	33.72	32.60	**31.93**	32.68	33.19	33.19	32.07	31.96
TBL	49.61	44.38	45.19	44.72	41.65	**40.94**	44.74	46.81	45.19	43.09	41.52
						CRPS skill score [%]					
BON	52.68	58.20	57.17	54.86	59.96	60.07	59.12	58.12	57.56	60.39	**60.87**
DRA	37.67	42.04	36.54	37.39	**44.17**	42.92	41.30	34.03	35.15	42.48	43.46
FPK	35.68	57.53	56.28	52.49	59.01	58.50	57.25	55.75	55.49	58.95	**59.70**
GWN	61.30	63.70	63.38	62.84	63.22	**65.85**	64.01	63.17	63.15	65.28	65.82
PSU	56.92	58.85	59.13	59.10	60.93	61.36	59.65	59.47	59.28	61.36	**61.53**
SXF	59.32	64.26	64.07	63.67	64.87	**65.60**	64.79	64.24	64.24	65.44	65.56
TBL	46.41	52.06	51.19	51.69	55.01	**55.78**	51.67	49.44	51.19	53.45	55.14

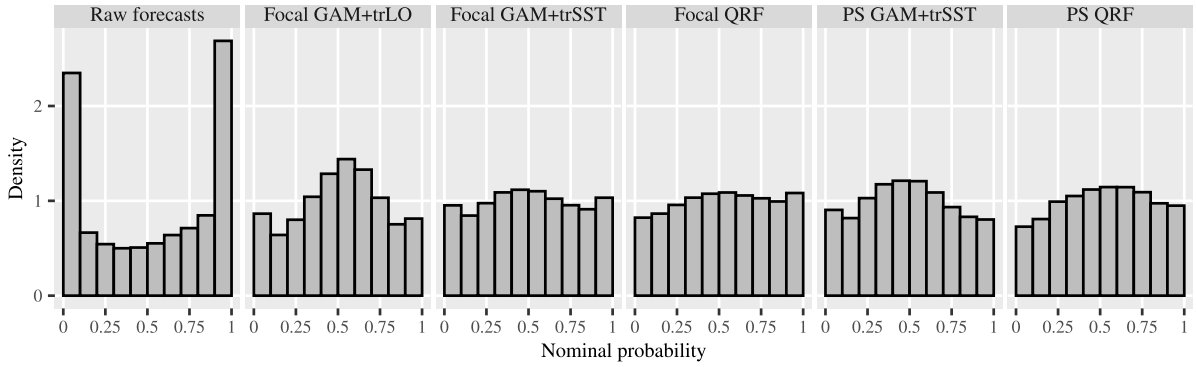

Fig. 4. Probability integral transform histogram of raw ensemble forecasts, focal and spatial post-processing techniques for all stations.

or better than spatial post-processing. This is the most notable result of this study. Also, in terms of CRPS and CRPS skill score, the GAM+trSST model is found mostly more accurate among parametric post-processing models in both focal and spatial post-processing. However, QRF model mostly outperformed parametric models in both focal and spatial post-processing. The performance of the reference LQR model against others in spatial post-processing is similar to that of in focal post-processing.

The prediction interval generated by the QRF model has PICP values higher than nominal coverage for spatial post-processing. Such high coverage actually is not desired as the predictive distribution is over-dispersed (or under-confident), where the grid operators may not get useful uncertainty information for decision-making. However, it is one of the highest quality post-processing models in terms of CRPS and CRPS skill score. Other post-processing techniques, both focal and spatial, are able to calibrate the predictive distributions to acceptable coverage levels.

Besides quantitative metrics, qualitative assessment tools are useful to evaluate the quality of predictive distributions. Probability integral transform (PIT) histogram is one of the commonly used qualitative tools to evaluate the calibration of probabilistic forecasts. A perfectly calibrated predictive distribution is expected to have a uniform PIT histogram. A U-shaped PIT histogram indicates that the predictive distribution is over-confident (or under-dispersed), whereas an inverse-U-shaped PIT histogram indicates under-confidence (or over-dispersion) [72]. Fig. 4 shows the PIT histogram of raw ensemble forecasts and post-processed forecasts. For visual clarity, forecasts from 7 stations are used together during plotting. Raw forecasts indicate strong under-dispersion as also evidenced by a U-shaped PIT histogram with large peaks at the tails. The histogram of the focal GAM+trLO model has a strong over-dispersion at the center, which is expected due to the shape of the truncated LO distribution. Based on the visual inspection of Fig. 4, it is clear that raw forecasts and the GAM+trLO model are not well calibrated, whereas GAM+trSST and QRF post-processing models have quasi-uniform PIT histograms. The focal GAM+trSST and QRF models are found slightly better dispersed than their spatial counterparts. Thus, focal models can be considered more calibrated (better dispersed) than spatial models based on PIT histograms.

A sharpness diagram is another graphical assessment tool that evaluates the sharpness of probabilistic forecasts. It illustrates the distribution of prediction interval width between upper and lower bounds of forecasts at each time t, i.e., $U_t - L_t$, which are defined in accordance with a central prediction interval. A sharp predictive distribution is expected to have low width. Fig. 5 shows the sharpness diagram of raw ensemble forecasts and post-processed forecasts. The upper and lower bounds of a predictive distribution with a nominal probability of 95% at time t are represented by U_t and L_t, respectively. Note

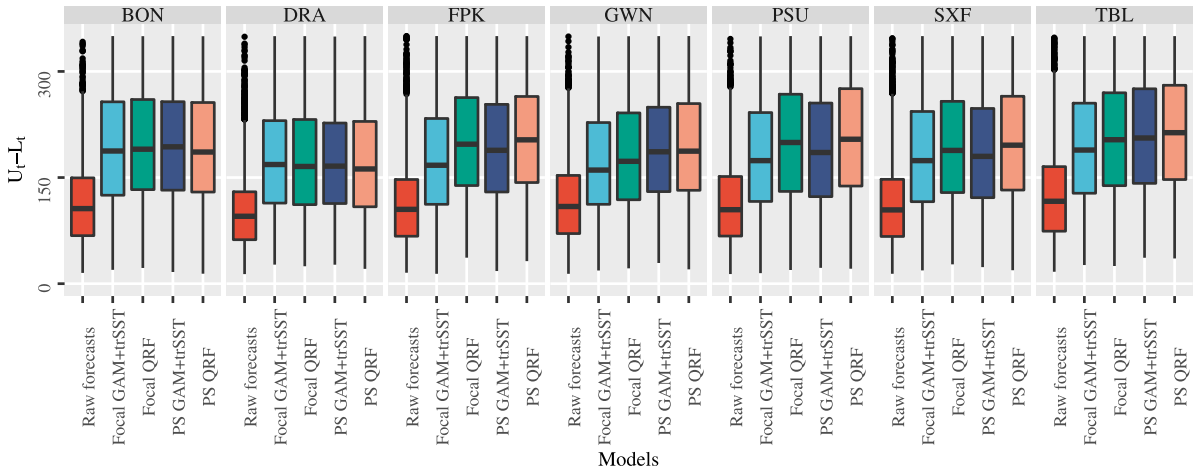

Fig. 5. Sharpness diagram of raw ensemble forecasts, focal and spatial post-processing techniques for all stations.

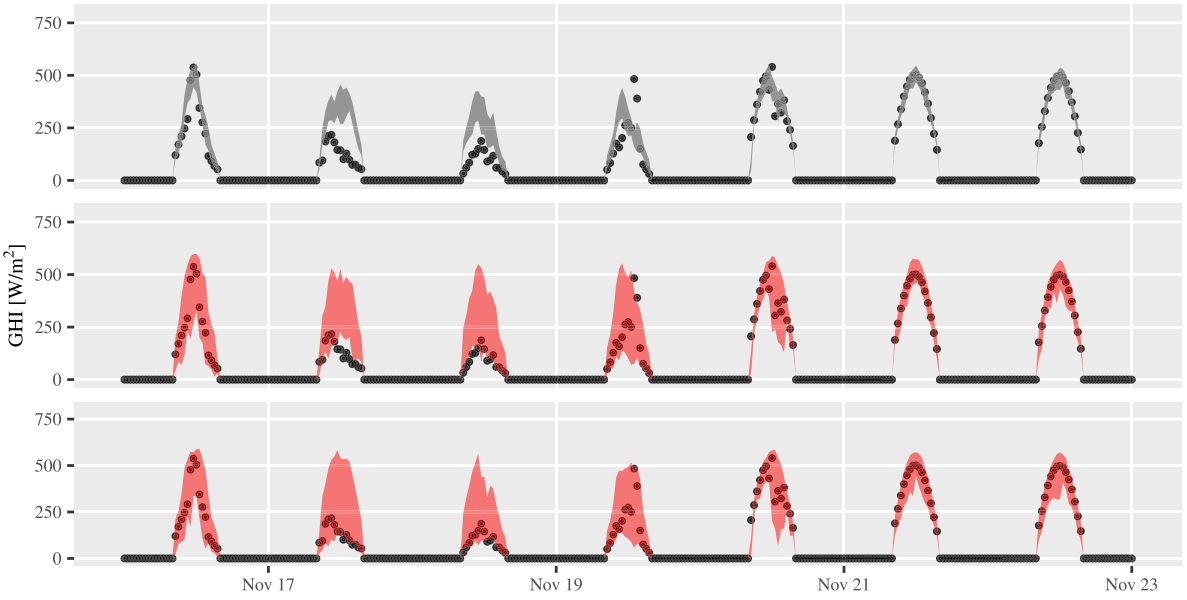

Fig. 6. An interval of GHI observations from a randomly selected week at BON station plotted together with the 30-min-ahead predictive density of raw ensemble forecasts (top), 30-min-ahead predictive density of focal (middle) and spatial (bottom) post-processed ensemble forecasts with GAM+trSST.

that the time instances where $U_t - L_t > 350$ W/m^2 are removed from the plot for visual clarity. Raw ensemble forecasts are mostly the sharpest. However, they are lacking in terms of calibration. The median sharpness (the center line in the box plot) of predictive intervals generated by focal and spatial post-processing models is about the same, where the focal models are found slightly sharper. Focal and spatial QRF models tend to generate large prediction interval widths in some forecast instances, which are undesired as such large intervals may not be informative to forecast users.

Fig. 6 shows the time series plot of GHI observations for a randomly selected week at BON station, together with the 30-min-ahead predictive density of raw ensemble forecasts, focal post-processing and spatial post-processing with GAM+trSST model for a 95% nominal coverage. Under-dispersion of raw ensemble forecasts is obvious, since many of the observations are not contained by the 95% prediction interval, whereas focal and spatial post-processing models are better dispersed. Though their predictive performance is quite similar, the density distribution of spatial post-processing seems more desirable. Based on the visual inspection of the time series plot, it is suggested to post-process raw ensemble forecasts generated by the dropout NN model to further improve dispersion; even using focal information only significantly improves the quality of prediction interval.

These results are in line with the literature of ensemble post-processing of other weather parameters. Post-processing using only focal forecasts is generally found better than using focal and spatial forecasts together unless a preselection is performed in spatial post-processing [39–41]. Recall that correlation-based preselection is applied before spatial post-processing. Cluster-based or analog-based methods have been proposed in the literature for other atmospheric variables, see, e.g., [39,40]. There are, of course, potential improvements that could stem from better preselection. However, we defer testing such methods to future studies. As evidenced by quantitative and qualitative assessment tools, raw ensemble forecasts generated by the dropout NN model need further adjustments, and by doing so, the quality of ensemble forecasts is significantly improved.

Satellite-derived irradiance is often a "go-to" resource to obtain regional irradiance information—it is technically not feasible to install and maintain sensor networks over large areas. However, image-to-irradiance conversion requires extensive physical modeling such as radiative transfer, which is associated with a steep learning curve, limiting its uptake among forecast practitioners. Considering the dropout

NN model substantially reduces the amount of physical modeling and the benefits offered by spatial post-processing make the techniques used in this study more appropriate for short-term areal forecasting applications. Also, the proposed techniques in this study have world-wide applicability, for satellite-derived irradiance data are available worldwide.

Despite the considerable progress in the development of irradiance ensemble post-processing, there are still many challenges to overcome. To the best of the authors' knowledge, the proposed ensemble post-processing techniques in the literature, as well as in this study, require ground-based sensor measurements and large datasets. This introduces challenges in operational implementation, such as, having proper quality control and verification of the data, handling a large amount of observation and forecast data, and following forecast submission requirements—just to name a few. These challenges remain largely unaddressed in the literature, which offers several avenues for future research. At the moment, a few works have considered replacing ground-based observations with satellite-derived ones for various solar energy meteorology applications include [18,46,47], which serve as background information, should one wish to move into this topic.

5. Conclusion

Data-driven ensemble solar forecasts are often found to be biased and/or under-dispersed. Such forecasts may lead to suboptimal decisions during grid operations. In this regard, post-processing is necessary to improve the calibration of these forecasts, e.g., by correcting the ensemble spread. Given the fact that solar forecasting is a spatio-temporal process, incorporating spatio-temporal information is likely to lead to high-accuracy forecasts. However, in ensemble post-processing, the potential benefits offered by spatio-temporal information have not been formally investigated. On that account, this article focuses on studying the potential benefits offered by incorporating spatio-temporal information into post-processing of ensemble solar forecasts.

More specifically, 30-min-ahead clear-sky index ensemble forecasts are post-processed using the forecasts from the neighboring satellite pixels around the location of interest. Spatial ensemble forecasts of neighboring satellite pixels are generated using a dropout neural network (NN) model with Monte Carlo sampling. Subsequently, various parametric and nonparametric post-processing techniques, such as non-homogenous regression, generalized additive model (GAM), or quantile regression, are applied to further improve the quality of NN-based ensemble forecasts. The models are trained using ground-based observations and satellite-derived irradiance data collected from research-grade data products, the SURFRAD and the NSRDB. Since incorporating spatial ensemble forecasts into post-processing results in a very large ensemble model pool, which is restricting computational performance, a naïve correlation-based preselection method is applied to narrow down the model pool. Popular quantitative and qualitative assessment tools are employed to evaluate the sharpness and calibration of the predictive distribution.

Parametric and nonparametric post-processing techniques significantly improved the quality of raw ensemble forecasts as evidenced by various metrics. Among parametric post-processing models, GAM with truncated skewed Student's t distribution consistently beats others, even nonparametric techniques in some cases. Nonparametric post-processing techniques mostly outperformed their parametric counterparts with and without forecast information from the neighboring satellite pixels. Incorporating spatial forecasts into post-processing has resulted in improved calibration and sharpness for some locations but not others. Hence, as in the case of most forecasting studies, the performance and the realized benefits are case dependent.

Although spatio-temporal information, as we have shown, may offer added benefits to post-processing, there are still several avenues for future research. Firstly, spatial post-processing can be computationally demanding if the number of ensemble members is large. This may introduce challenges in operational implementation. Hence, employing a preselection method, to reduce the number of component forecasts to be post-processed, is to be explored. Secondly, spatial ensemble forecasting should be studied using realistically distributed satellite pixels over a region considering locations of electrical substations and photovoltaic plants as opposed to a "tidy" selection as in this study.

CRediT authorship contribution statement

Gokhan Mert Yagli: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization. **Dazhi Yang:** Methodology, Validation, Writing – review & editing. **Dipti Srinivasan:** Validation, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research is supported by the National Research Foundation Singapore under EMA-ESG Singapore grant (Award Number: [NRF2019NRF-CG002-004]).

References

[1] Wilks DS. Statistical methods in the atmospheric sciences, Vol. 100. Elsevier; 2019, http://dx.doi.org/10.1016/C2017-0-03921-6.
[2] Yang D. A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES). J Renew Sustain Energy 2019;11(2):022701. http://dx.doi.org/10.1063/1.5087462.
[3] Hong T, Pinson P, Wang Y, Weron R, Yang D, Zareipour H. Energy forecasting: A review and outlook. IEEE Open Access J Power Energy 2020;7:376–88. http://dx.doi.org/10.1109/OAJPE.2020.3029979.
[4] Ren Y, Suganthan PN, Srikanth N. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. Renew Sustain Energy Rev 2015;50:82–91. http://dx.doi.org/10.1016/j.rser.2015.04.081.
[5] Yang D, van der Meer D. Post-processing in solar forecasting: Ten overarching thinking tools. Renew Sustain Energy Rev 2021;140:110735. http://dx.doi.org/10.1016/j.rser.2021.110735.
[6] Yagli GM, Yang D, Srinivasan D. Ensemble solar forecasting using data-driven models with probabilistic post-processing through GAMLSS. Sol Energy 2020;208:612–22. http://dx.doi.org/10.1016/j.solener.2020.07.040.
[7] Doubleday K, Jascourt S, Kleiber W, Hodge B-M. Probabilistic solar power forecasting using Bayesian model averaging. IEEE Trans Sustain Energy 2021;12(1):325–37.
[8] Bakker K, Whan K, Knap W, Schmeits M. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. Sol Energy 2019;191:138–50. http://dx.doi.org/10.1016/j.solener.2019.08.044.
[9] Yagli GM, Yang D, Srinivasan D. Reconciling solar forecasts: Sequential reconciliation. Sol Energy 2019;179:391–7. http://dx.doi.org/10.1016/j.solener.2018.12.075.
[10] Yang D. Reconciling solar forecasts: Probabilistic forecast reconciliation in a nonparametric framework. Sol Energy 2020;210:49–58. http://dx.doi.org/10.1016/j.solener.2020.03.095, Special Issue on Grid Integration.
[11] Yagli GM, Yang D, Srinivasan D. Reconciling solar forecasts: Probabilistic forecasting with homoscedastic Gaussian errors on a geographical hierarchy. Sol Energy 2020;210:59–67. http://dx.doi.org/10.1016/j.solener.2020.06.005, Special Issue on Grid Integration.
[12] Vannitsem S, Wilks DS, Messner J. Statistical postprocessing of ensemble forecasts. Elsevier; 2018, http://dx.doi.org/10.1016/C2016-0-03244-8.
[13] Huang G, Li Z, Li X, Liang S, Yang K, Wang D, et al. Estimating surface solar irradiance from satellites: Past, present, and future perspectives. Remote Sens Environ 2019;233:111371. http://dx.doi.org/10.1016/j.rse.2019.111371.
[14] Yang D, Kleissl J, Gueymard CA, Pedro HTC, Coimbra CFM. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Sol Energy 2018;168:60–101. http://dx.doi.org/10.1016/j.solener.2017.11.023, Advances in Solar Resource Assessment and Forecasting.
[15] Kumar DS, Yagli GM, Kashyap M, Srinivasan D. Solar irradiance resource and forecasting: a comprehensive review. IET Renew Power Gener 2020;14(10):1641–56. http://dx.doi.org/10.1049/iet-rpg.2019.1227.

[16] van der Meer D, Widén J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew Sustain Energy Rev 2018;81:1484–512. http://dx.doi.org/10.1016/j.rser.2017.05.212.

[17] Buizza R. Introduction to the special issue on "25 years of ensemble forecasting". Q J R Meteorol Soc 2019;145(S1):1–11. http://dx.doi.org/10.1002/qj.3370.

[18] Yagli GM, Yang D, Gandhi O, Srinivasan D. Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance? Appl Energy 2020;259:114122. http://dx.doi.org/10.1016/j.apenergy.2019.114122.

[19] Yang D, Dong Z. Operational photovoltaics power forecasting using seasonal time series ensemble. Sol Energy 2018;166:529–41. http://dx.doi.org/10.1016/j.solener.2018.02.011.

[20] Yagli GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019;105:487–98. http://dx.doi.org/10.1016/j.rser.2019.02.006.

[21] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Proceedings of the 33rd international conference on international conference on machine learning , Vol. 48. ICML'16, JMLR.org; 2016, p. 1050–9, URL https://dl.acm.org/doi/10.5555/3045390.3045502.

[22] Lee H, Lee B-T. Confidence-aware deep learning forecasting system for daily solar irradiance. IET Renew Power Gener 2019;13(10):1681–9. http://dx.doi.org/10.1049/iet-rpg.2018.5354.

[23] Lee H, Kim N-W, Lee J-G, Lee B-T. Uncertainty-aware forecast interval for hourly PV power output. IET Renew Power Gener 2019;13(14):2656–64. http://dx.doi.org/10.1049/iet-rpg.2019.0300.

[24] de Jongh S, Riedel T, Mueller F, Yacoub AE, Suriyah M, Leibfried T. Spatio-temporal short term photovoltaic generation forecasting with uncertainty estimates using machine learning methods. In: 2020 55th international universities power engineering conference. 2020, p. 1–6. http://dx.doi.org/10.1109/UPEC49904.2020.9209764.

[25] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929–58.

[26] Sengupta M, Habte A, Wilbert S, Gueymard C, Remund J. Best practices handbook for the collection and use of solar resource data for solar energy applications. Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States); 2021.

[27] Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. J R Stat Soc Ser B Stat Methodol 2007;69(2):243–68. http://dx.doi.org/10.1111/j.1467-9868.2007.00587.x.

[28] Miller SD, Rogers MA, Haynes JM, Sengupta M, Heidinger AK. Short-term solar irradiance forecasting via satellite/model coupling. Sol Energy 2018;168:102–17. http://dx.doi.org/10.1016/j.solener.2017.11.049, Advances in Solar Resource Assessment and Forecasting.

[29] Yang D. Validation of the 5-min irradiance from the National Solar Radiation Database (NSRDB). J Renew Sustain Energy 2021;13(1):016101. http://dx.doi.org/10.1063/5.0030992.

[30] Beucler T, Pritchard M, Rasp S, Ott J, Baldi P, Gentine P. Enforcing analytic constraints in neural networks emulating physical systems. Phys Rev Lett 2021;126:098302. http://dx.doi.org/10.1103/PhysRevLett.126.098302.

[31] Rasp S, Dueben PD, Scher S, Weyn JA, Mouatadid S, Thuerey N. WeatherBench: A benchmark data set for data-driven weather forecasting. J Adv Modelling Earth Syst 2020;12(11). http://dx.doi.org/10.1029/2020MS002203.

[32] Yang D, Gu C, Dong Z, Jirutitijaroen P, Chen N, Walsh WM. Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging. Renew Energy 2013;60:235–45. http://dx.doi.org/10.1016/j.renene.2013.05.030.

[33] Yang D, Dong Z, Reindl T, Jirutitijaroen P, Walsh WM. Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage. Sol Energy 2014;103:550–62. http://dx.doi.org/10.1016/j.solener.2014.01.024.

[34] Yang D. Ultra-fast preselection in lasso-type spatio-temporal solar forecasting problems. Sol Energy 2018;176:788–96. http://dx.doi.org/10.1016/j.solener.2018.08.041.

[35] Yang D. Spatial prediction using kriging ensemble. Sol Energy 2018;171:977–82. http://dx.doi.org/10.1016/j.solener.2018.06.105.

[36] Yagli GM, Tay JWE, Yang D. Ensemble kriging for environmental spatial processes. In: 2019 IEEE international conference on big data. 2019, p. 3947–50. http://dx.doi.org/10.1109/BigData47090.2019.9005731.

[37] Grönquist P, Yao C, Ben-Nun T, Dryden N, Dueben P, Li S, et al. Deep learning for post-processing ensemble weather forecasts. Phil Trans R Soc A 2021;379(2194):20200092. http://dx.doi.org/10.1098/rsta.2020.0092.

[38] Vannitsem S, Bremnes JB, Demaeyer J, Evans GR, Flowerdew J, Hemri S, Lerch S, Roberts N, Theis S, Atencia A, Bouallègue ZB, Bhend J, Dabernig M, Cruz LD, Hieta L, Mestre O, Moret L, Plenković IO, Schmeits M, Taillardat M, den Bergh JV, Schaeybroeck BV, Whan K, Ylhaisi J. Statistical postprocessing for weather forecasts: Review, challenges, and avenues in a big data world. Bull Am Meteorol Soc 2021;102(3):E681–99. http://dx.doi.org/10.1175/BAMS-D-19-0308.1.

[39] Lerch S, Baran S. Similarity-based semilocal estimation of post-processing models. J R Stat Soc Ser C Appl Stat 2017;66(1):29–51. http://dx.doi.org/10.1111/rssc.12153.

[40] Junk C, Delle Monache L, Alessandrini S. Analog-based ensemble model output statistics. Mon Weather Rev 2015;143(7):2909–17. http://dx.doi.org/10.1175/MWR-D-15-0095.1.

[41] Rasp S, Lerch S. Neural networks for postprocessing ensemble weather forecasts. Mon Weather Rev 2018;146(11):3885–900. http://dx.doi.org/10.1175/MWR-D-18-0187.1.

[42] Augustine JA, DeLuisi JJ, Long CN. SURFRAD—A National surface radiation budget network for atmospheric research. Bull Am Meteorol Soc 2000;81(10):2341–58. http://dx.doi.org/10.1175/1520-0477(2000)081<2341:SANSRB>2.3.CO;2.

[43] Yang D. SolarData: An R package for easy access of publicly available solar datasets. Sol Energy 2018;171:A3–12. http://dx.doi.org/10.1016/j.solener.2018.06.107.

[44] Yang D. SolarData package update v1.1: R functions for easy access of Baseline Surface Radiation Network (BSRN). Sol Energy 2019;188:970–5. http://dx.doi.org/10.1016/j.solener.2019.05.068.

[45] Long CN, Dutton EG. BSRN global network recommended QC tests, V2. x. PANGAEA; 2010, doi: 10013/epic.38770.d001.

[46] Yang D, Perez R. Can we gauge forecasts using satellite-derived solar irradiance? J Renew Sustain Energy 2019;11(2):023704. http://dx.doi.org/10.1063/1.5087588.

[47] Yang D. Post-processing of NWP forecasts using ground or satellite-derived data through kernel conditional density estimation. J Renew Sustain Energy 2019;11(2):026101. http://dx.doi.org/10.1063/1.5088721.

[48] Sengupta M, Habte A, Wilbert S, Gueymard C, Remund J. Best practices handbook for the collection and use of solar resource data for solar energy applications. Tech. rep., 3rd ed.. 2021, http://dx.doi.org/10.2172/1778700.

[49] Yang D, Bright JM. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years. Sol Energy 2020;210:3–19. http://dx.doi.org/10.1016/j.solener.2020.04.016, Special Issue on Grid Integration.

[50] Yang D. A correct validation of the National Solar Radiation Data Base (NSRDB). Renew Sustain Energy Rev 2018;97:152–5. http://dx.doi.org/10.1016/j.rser.2018.08.023.

[51] Sengupta M, Xie Y, Lopez A, Habte A, Maclaurin G, Shelby J. The National Solar Radiation Data Base (NSRDB). Renew Sustain Energy Rev 2018;89:51–60. http://dx.doi.org/10.1016/j.rser.2018.03.003.

[52] Gueymard CA. REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation — validation with a benchmark dataset. Sol Energy 2008;82(3):272–85. http://dx.doi.org/10.1016/j.solener.2007.04.008.

[53] Sun X, Bright JM, Gueymard CA, Acord B, Wang P, Engerer NA. Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis. Renew Sustain Energy Rev 2019;111:550–70. http://dx.doi.org/10.1016/j.rser.2019.04.006.

[54] Sun X, Bright JM, Gueymard CA, Bai X, Acord B, Wang P. Worldwide performance assessment of 95 direct and diffuse clear-sky irradiance models using principal component analysis. Renew Sustain Energy Rev 2021;135:110087. http://dx.doi.org/10.1016/j.rser.2020.110087.

[55] Yang D. Choice of clear-sky model in solar forecasting. J Renew Sustain Energy 2020;12(2):026101. http://dx.doi.org/10.1063/5.0003495.

[56] Yang D, Alessandrini S, Antonanzas J, Antonanzas-Torres F, Badescu V, Beyer HG, Blaga R, Boland J, Bright JM, Coimbra CFM, David M, Frimane Â, Gueymard CA, Hong T, Kay MJ, Killinger S, Kleissl J, Lauret P, Lorenz E, van der Meer D, Paulescu M, Perez R, Perpiñán-Lamigueiro O, Peters IM, Reikard G, Renné D, Saint-Drenan Y-M, Shuai Y, Urraca R, Verbois H, Vignola F, Voyant C, Zhang J. Verification of deterministic solar forecasts. Sol Energy 2020;210:20–37. http://dx.doi.org/10.1016/j.solener.2020.04.019, Special Issue on Grid Integration.

[57] Hyndman R, Athanasopoulos G. Forecasting: Principles and practice. Australia, OTexts: Melbourne; 2021, https://www.otexts.com/fpp3.

[58] Gal Y, Hron J, Kendall A. Concrete dropout. In: Advances in neural information processing systems 30: Annual conference on neural information processing systems 2017. 2017, p. 3581–90, URL https://dl.acm.org/doi/10.5555/3294996.3295116.

[59] Keydana S. RStudio AI blog: You sure? A Bayesian approach to obtaining uncertainty estimates from neural networks. 2018, URL https://blogs.rstudio.com/tensorflow/posts/2018-11-12-uncertainty_estimates_dropout/.

[60] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016, https://www.deeplearningbook.org.

[61] Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? In: Proceedings of the 31st international conference on neural information processing systems. Red Hook, NY, USA: Curran Associates Inc.; 2017, p. 5580—5590, URL https://dl.acm.org/doi/10.5555/3295222.3295309.

[62] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: 3rd international conference on learning representations. 2015, URL http://arxiv.org/abs/1412.6980, 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.

[63] Gneiting T, Raftery AE, Westveld AH, Goldman T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon Weather Rev 2005;133(5):1098–118. http://dx.doi.org/10.1175/MWR2904.1.

[64] Scheuerer M, Möller D. Probabilistic wind speed forecasting on a grid based on ensemble model output statistics. Ann Appl Stat 2015;9(3):1328–49. http://dx.doi.org/10.1214/15-AOAS843.
[65] Rigby RA, Stasinopoulos DM, Heller GZ, De Bastiani F. Distributions for modeling location, scale, and shape: Using GAMLSS in R. CRC Press; 2019.
[66] Gebetsberger M, Messner JW, Mayr GJ, Zeileis A. Estimation methods for nonhomogeneous regression models: Minimum continuous ranked probability score versus maximum likelihood. Mon Weather Rev 2018;146(12):4323–38. http://dx.doi.org/10.1175/MWR-D-17-0364.1.
[67] Stasinopoulos DM, Rigby RA, Heller GZ, Voudouris V, De Bastiani F. Flexible regression and smoothing: using GAMLSS in R. Chapman and Hall/CRC; 2017.
[68] Hastie TJ, Tibshirani RJ. Generalized additive models, Vol. 43. CRC Press; 1990.
[69] Koenker R. Quantile regression. Econometric society monographs, Cambridge University Press; 2005, http://dx.doi.org/10.1017/CBO9780511754098.
[70] Meinshausen N. Quantile regression forests. J Mach Learn Res 2006;7:983–99, URL http://jmlr.org/papers/v7/meinshausen06a.html.
[71] Breiman L. Random forests. Mach Learn 2001;45(1):5–32. http://dx.doi.org/10.1023/A:1010933404324.
[72] Gneiting T, Katzfuss M. Probabilistic forecasting. Annu Rev Stat Appl 2014;1(1):125–51. http://dx.doi.org/10.1146/annurev-statistics-062713-085831.
[73] Lauret P, David M, Pinson P. Verification of solar irradiance probabilistic forecasts. Sol Energy 2019;194:254–71. http://dx.doi.org/10.1016/j.solener.2019.10.041.
[74] Yang D. A universal benchmarking method for probabilistic solar irradiance forecasting. Sol Energy 2019;184:410–6. http://dx.doi.org/10.1016/j.solener.2019.04.018.